eDRN Catalog and Archive Service (eCAS)
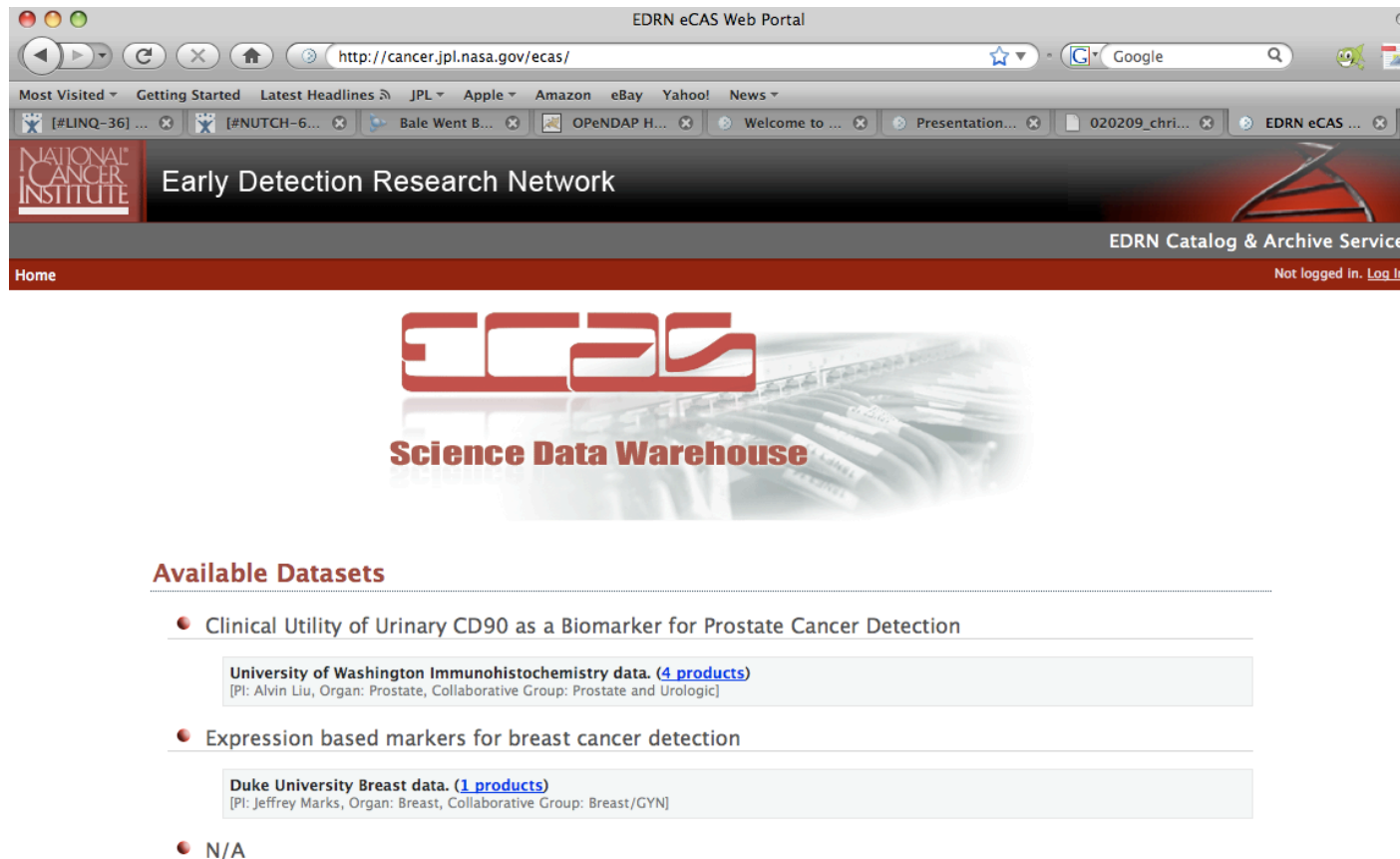
Chris A. Mattmann

Outline

- Interface/Access
- Current Release and Planned Updates
- Security
- Release Policies
- Data Sharing

eCAS Interface: Splash Page

- ## http://ecas.jpl.nasa.gov
  - Operational 0.3 beta release

eCAS Interface: Logged in Dataset View

eCAS Interface: Logged in Product View

Capabilities

- Current (as of 0.3.0 beta)
  - Download datasets as zip files
    - Zips include zips of products (and their metadata), as well as dataset-level metadata
  - Download products as zip files
  - Browse of dataset and product metadata
  - Dump RDF dataset and product information into EKE
- Planned (in next 2 months, 2-4 releases)
  - Improved browsing
    - pagination of results
  - Curation interface
    - Talk more about this later, with John's slides
  - Bug fixes/maintenance
    - Dataset delivery and proper cleanup

Current Datasets

- SELDI Phase I and II,MALDI Dilution Data, Miscellaneous PI data, (Immunohistochemistry, LabMAP Mass Spectrometry, FISH MSA data)
- Test data from Bill Grizzle (as of 0.3.0 beta release)
  - EGFR Translocation Data
  - Pre-invasive Neoplasia Image and Summary Data
- Sites
  - NIST, FHCRC, EVMS, UAB, Duke, Univ Pitt, Univ Wash

Security

- Current
  - Integration with EDRN Informatics Center LDAP server
  - Basic authentication, to guard datasets and products, and download
- Planned
  - Granular authorization, based on roles (Heather's talk)
    - Need to develop dataset, product-level policy, as well as metadata attribute browse, and downloading policy

Release Policies

- Software
  - Documented at:
    https://cancer.jpl.nasa.gov/documents/applications/catalog-and-archive-service/deploying-the-ecas-to-operations
    - Describes how to deploy the eCAS software to operations
- Datasets
  - As datasets are acquired from PIs (via mobile eCAS, or via WebDAV, or via CD/DVD), there is a vetting process involving curation activities (will be discussed in more detail by John later)
    - Dataset-level Metadata curation
    - Metadata extraction for Product-level curation
    - Bulk Ingest, or product-by-product
    - Validation/Verification
      - Performed on tumor (dev) first, then performed on cancer (ops)
  - Datasets released currently along with software, but can be a separate activity as well

Data Sharing

- Defined in part by security policies and authorization at the eCAS/eCAS curation level
- Defined also in part through RDF dump of product/dataset metadata to portal and ability to download datasets
- Requests for eCAS data from private and research communities
  - Data Disclaimer
- eCAS meant to be a software/data enabler, not a lawyer

Questions